Violence and Incitement.

Jorge Mauricio Rey Molina/CCMS 4779035

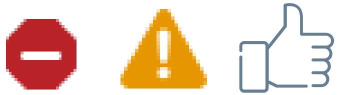
Lesson plan.

01 **Introduction/Warm Up** (Disclaimer, Objectives & Policy Rationale)

02 **Presentation** (What we remove?)

03 **Practice** (Jobs examples)

04 **Conclusion**



Introduction



Disclaimer



Remember that all intellectual content was taken from the official Meta page, in its Transparency Center section, all rights and information are reserved to them.

REMEMBER, this presentation may have content that is not suitable for everyone, if you have any problems, please raise your hand, for your health we will act according to our well-being protocol.

Objectives

General

That the trainee understands and know how to act the concept of violence and Incitement policy created by Meta.

Specific 1

That the trainee understands what a violent threats is and how to identify it.

Specific 2

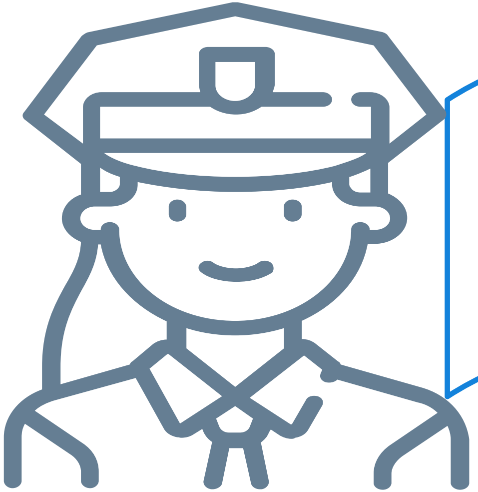
That the trainee identifies the different targets which Meta aims to protect.

01

Policy Rationale



“We aim to prevent potential offline violence that may be related to content on our platforms”



We remove language that incites or facilitates violence and credible threats to public or personal safety.

02 We Remove:



Take into account

"We Eliminate threats of violence against various targets. Threats can be expressed in various ways, such as statements of intent, calls to action, or aspirational statements"





Universal protections for everyone

Everyone is protected from the following threats:

Violence Types

High violence

•Threats of violence that could lead to death (or other forms of high-severity violence)

Threats of violence that could lead to serious injury (mid-severity violence).

Mid violence

Low violence

Private adults (when self-reported), children and persons or groups of people targeted on the basis of their protected characteristic(s), are protected from threats of low-severity violence.

Other violence

In addition to all of the protections listed above, we remove the following:



- Offers or solicitation for high-severity violent services (e.g., hitmen, assassins)
- Instructions on weapon-making or usage explicitly aimed at serious harm or death, except in self-defense, military training, video games, or news.
- Explosive-making instructions, unless for non-violent purposes (e.g., games, education, fireworks).
- Threats to bring weapons or forcibly enter locations (e.g., worship places, schools, polling places) during heightened risk.
- Threats related to voting, registration, or election administration, regardless of target.

Actions to take.

Delete.

This action is taken when the content violates our violence and incitement policy.

NA

No Action.

This action is taken when there is no violation of our policy.

Escalate.

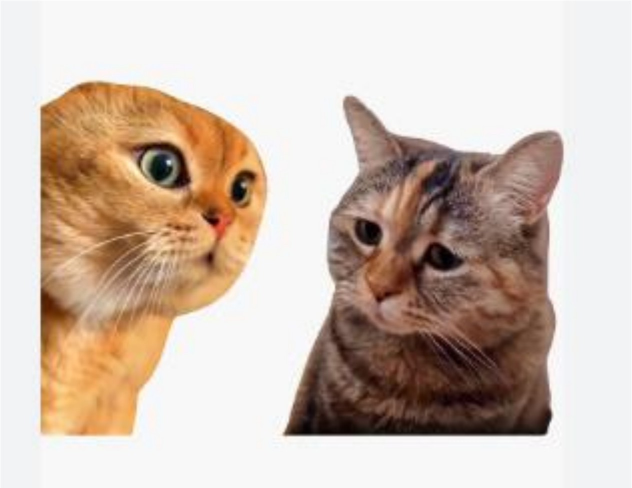
This action is taken when there is a report of high violence with a specific place and date in the future.

03 Examples:



Example #1

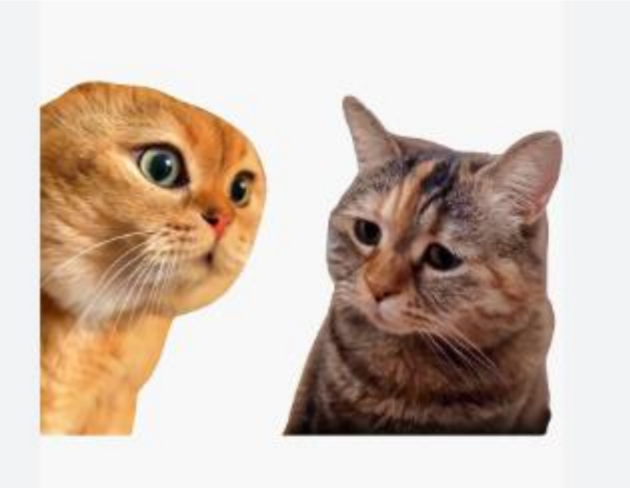
I'm going to shoot you with a gun



OH No!

Example #1

I'm going to shoot you with a gun

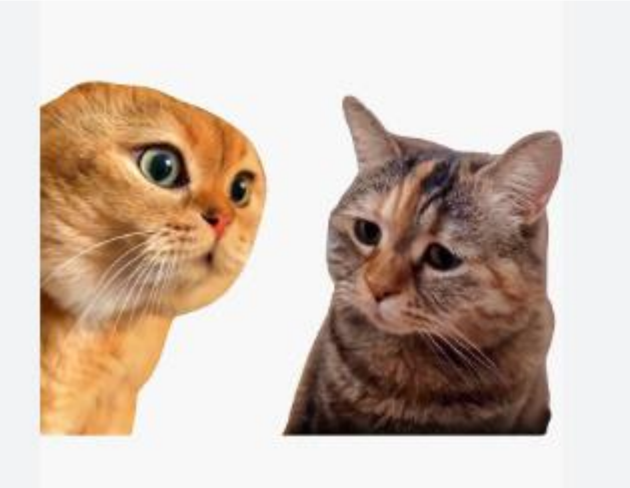


OH No!

Action: Delete High severity violence

Example #2

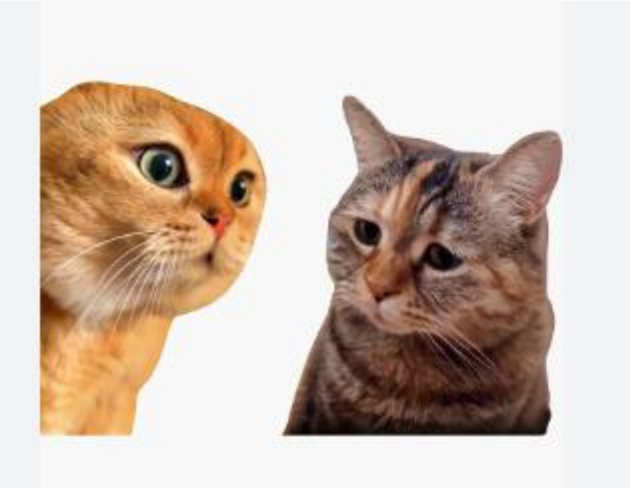
Tomorrow I will hit you at school.



OH No! Again...

Example #2

Tomorrow I will hit you at school.

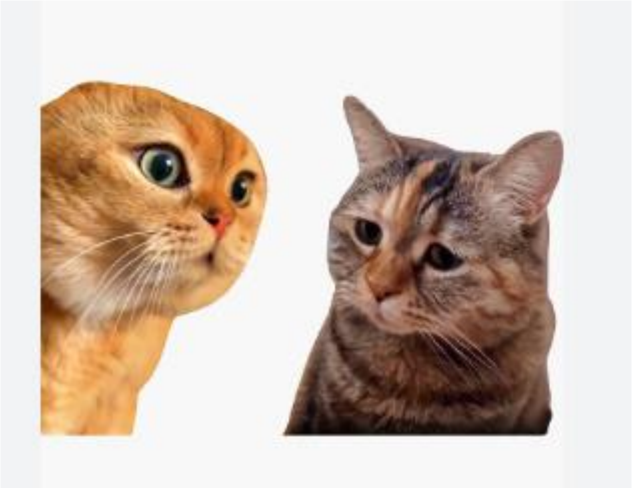


OH No! Again...

Action: Delete Mid severity violence

Example #3

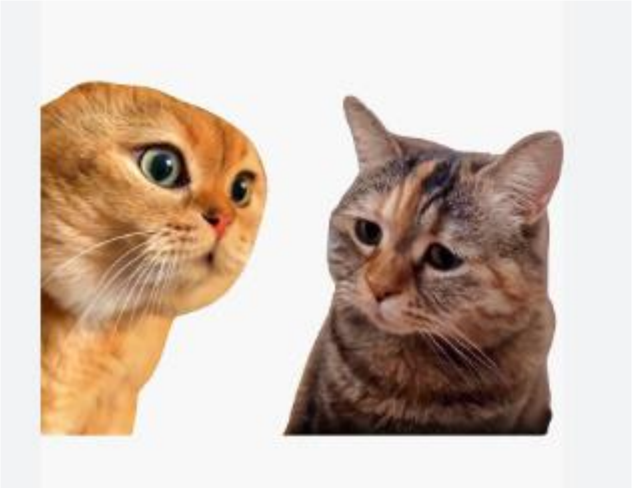
I love you!



OH No!

Example #2

Tomorrow I will hit you at school.



OH No! Again...

Action: No action

04 Conclusions:



Thanks!

Do you have any questions?